



Regression Modeling

An Appraiser's Perspective

Arizona Tax Conference, September 1, 2017

Presentation Outline and Format

Goals!

- Explain how to set your jurisdiction up for regression success
- Talk about the benefits of developing market values based on regression
- Generally describe the regression modeling process
- Identify appraisal touchpoints in the regression process
- Live Demo?

Characteristics of an Effective Assessment Office

Many of the characteristics of an effective assessment office have a direct impact on effective modeling:

- Legal Framework – disclosure laws, specialized appraisals
- Effective Valuation Techniques
- Periodic Ratio Studies
- Competent Staff - Effective Training
- Effective Internal Controls – Planning and QC
- Continuous Monitoring Activities
- Accurate Market Data

Essential Assessor Functions

Discover

- Normally related to identifying taxable property
- In this context, we mean a full population of properties



List

- Correctly listing the properties is vital to the reliability of your valuations, and the public perception of your office
- The inventory of a property becomes the basis of every valuation task we undertake



Value

- A full population of properties, correctly listed, provides the best basis for a correct valuation, using the three standard approaches



Three Approaches that Estimate Market Value

Part of our legal framework is that our valuations are synonymous with Market Value. We use **three recognized approaches**:

- **Cost** - Various Types (MLPLD, RCNLD) that rely heavily on a full inventory of the property
- **Income** - Multiple income approaches. Additional income data required, but also heavily reliant upon property inventory
- **Market** - Multiple market approaches. Can use the same inventory of property that cost uses, but may receive transformations on the data

All three approaches are calibrated by Sales

Important to have Multiple Approaches to Value

Although all property types do not lend themselves to all approaches to value, all approaches should be considered.

- Different approaches can be tested against sales to determine which produces the best value by property type
- Lends additional appraisal credibility to your office
- Other approaches can be used to support your determined approach.
- Sometimes you need to review the approaches just to determine which approach **NOT** to use

Inventory of Property is the basis of all good valuation

In addition to being intuitive, a **full, accurate** and **equitable** inventory is required in an effective assessment program

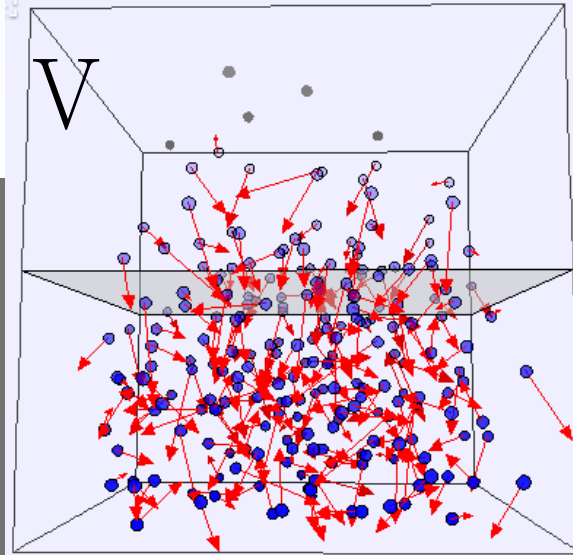
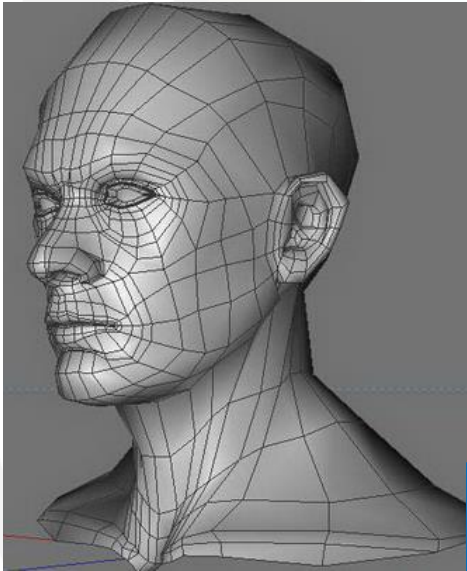
- **Full** – a comprehensive set of data that fully describes the physical and economic factors present in a property
- **Accurate** – data is frequently checked for quality, specifically as it relates to accurate representation of the property
- **Equitable** – The data itself should have been derived from standardized procedures so that data describing one thing on a property means the same thing on another property.

Cost approaches tend to be based on a full inventory of the property. In this way, a finely-tuned cost approach becomes a stable basis for other valuations.

Modeling Overview

Modeling is everywhere...

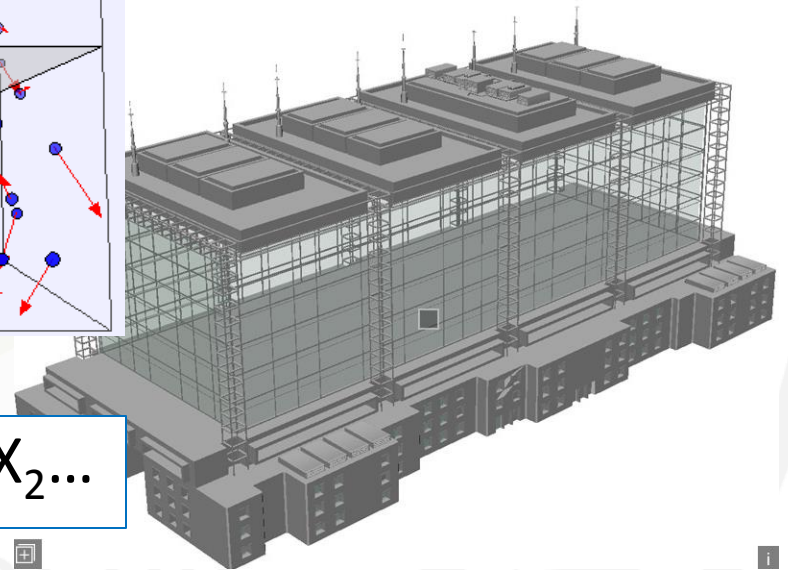
$$I = R \times V$$



$$S = b_0 + b_1X_1 + b_2X_2...$$

[View source image](#)

$$E = MC^2$$



Modeling defined: to make or construct a descriptive or representational model of something

Modeling Overview

Mathematical modeling can be done in various ways:

- Linear, Nonlinear, Static, Dynamic, Explicit, Implicit, Discrete, Deterministic, Stochastic, blah, blah, blah
- Linear FTW!
- Linear models are simplest to explain and can be easily calibrated
- However, linear models suffer in capturing non-linear trends

Regression Modeling Origination

“The use of regression models in statistical analysis was pioneered by (Sir) Francis Galton, a 19th Century scientist and explorer who might be considered a model for the Indiana Jones character of the movies. Early in his career, after he inherited a fortune and quit medical school, he went on two expeditions to Africa, the first to the upper Nile Valley and the second through parts of south-west Africa, sandwiched around 5 years of enjoying the sporting life. Based on what he learned from these adventures he wrote two best-selling books—The Art of Travel and its sequel, The Art of Rough Travel—which offered practical advice to future explorers on topics such as how to treat spear wounds and pull your horse out of quicksand, and he introduced a new item of camping gear to the Western world: the sleeping bag.”

- *Notes on linear regression analysis, Robert Nau, Fuqua School of Business, Duke University*

Regression Modeling as a Process to reach Market Value

- Regression modeling in our context attempts to predict what a property will sell for (Market Value) based on sales of other similar properties.
- This process determines the value of each characteristic influencing the market
- This statistical technique attempts to minimize the overall “error” in its predictions of what a property should sell for.

Pros and Cons of Regression Modeling

Pros:

- Provides an additional, widely recognized approach to value during initial valuation (market approach)
- Highly agile, and can directly represent current market dynamics.
- Products of the model can be used in other appraisal scenarios:
 - Use coefficients in appeals
 - Enhanced appraiser knowledge of what moves the market
 - Help determine appraisal priorities

Models are Highly Agile

Consider if you suddenly saw a new improvement come into your jurisdiction...

- *Your model could be updated that year, and a contributory value determined for the item even before M&S costs are established*

Consider if some existing type of construction suddenly became “Hot” ...

- *Your model can immediately capture the market value of the item by determining its market influence.*

Consider if some new influence entered into the marketplace – within a NBHD

- *The influence itself could be modeled without the need to further split the NBHD*

Pros and Cons of Regression Modeling

Cons:

- None!
- Can be difficult to explain to the public.
- Requires specialized knowledge in order to create good values.
- Can become *self-indulgent* to the modeler.
- Requires buy-in from appraisers.

Requirements of a Successful Regression program

- Ability of CAMA System to incorporate regression results
- Full buy-in from staff
- In-house knowledge or ability to contract the work
- Accurate population data, **including location data**
- Uniform processes to be used in developing and applying the model
- Sufficient, representative, accurate sales data.

Setting up for Success: Clean, Organized and Equitable Property Characteristic Data

“GIGO”

Development of a Quality Control program is essential

- As stated previously, your data is the beginning of a process that ends with a taxpayer opening their wallet and paying a tax bill. As such, it needs to be as good as we can make it
- Your CAMA system can either hamstring you or be your best friend

Clean, Organized and Equitable Property Characteristic Data

Quality Control of your data means verifying the accuracy of your data in three broad categories:

- Accuracy of Data Entry – **Easy**
 - CAMA System mainly responsible
 - Field Level Constraints – lookup lists
 - Edit Checks - Soft and hard
 - Accuracy of Property Inventory – **Harder**
 - Querying your data and looking for outliers
 - Looking for problems in converted data
 - Looking for newly arising issues, either systemic or procedural
 - Accuracy of Production Work to Established Standards – **Hardest?**
 - Performing analysis in qualitative decisions of staff
 - Reviewing field decisions against real world examples
 - Confirming that office policy and standard appraisal practices are being uniformly applied by all staff.
 - Confirming that production standards are being met while quotas are fulfilled
- 

The background of the slide features a large, light gray watermark of the Seal of Mohave County, Arizona. The seal is circular and contains a landscape with a mountain, a river, and a herd of cattle. The text "SEAL OF MOHAVE COUNTY ARIZONA" is written around the perimeter, and "DITAT DEUS" is at the bottom.

Questions so far?

A Quick Summary of the Modeling Process

- Definition of the Appraisal Problem
- Data Collection
- Property Data Exploration and Market Analysis
- Specification of the Model
- Sales Data Exploration
- Calibration of the Model
- Model Testing and Quality Control

Defining the Appraisal Problem

Successful Modeling requires a concrete understanding of what you are attempting to value.

- Your building inventory is your materials box, but can even then be lacking, depending on the appraisal problem.
- Different property types can lend themselves to different model specifications.
- Availability of sales also can constrain the type of models possible.
- Models with plenty of sales and homogeneous areas tend to be most successful.

Data Collection - Pulling Data From Sources

Yay Arizona! – We live in a state that requires Affidavits of Value!

- The determination of the appraisal problem will determine the data sources you will use.
- Your best friend – your cost data inventory
- Not so good friends but ones you might wave at in the store – MLS, Costar, property manager data

Data Exploration and Market Analysis - Data Review

"You will find only what you bring in"

- Yoda

A well-developed model will represent the population it values.

Review:

- Types of Construction
- Qualities
- General Amenities
- Ages
- Sizes
- Median Land and Improvement Values

Importance of good Neighborhoods

Location, location, location!

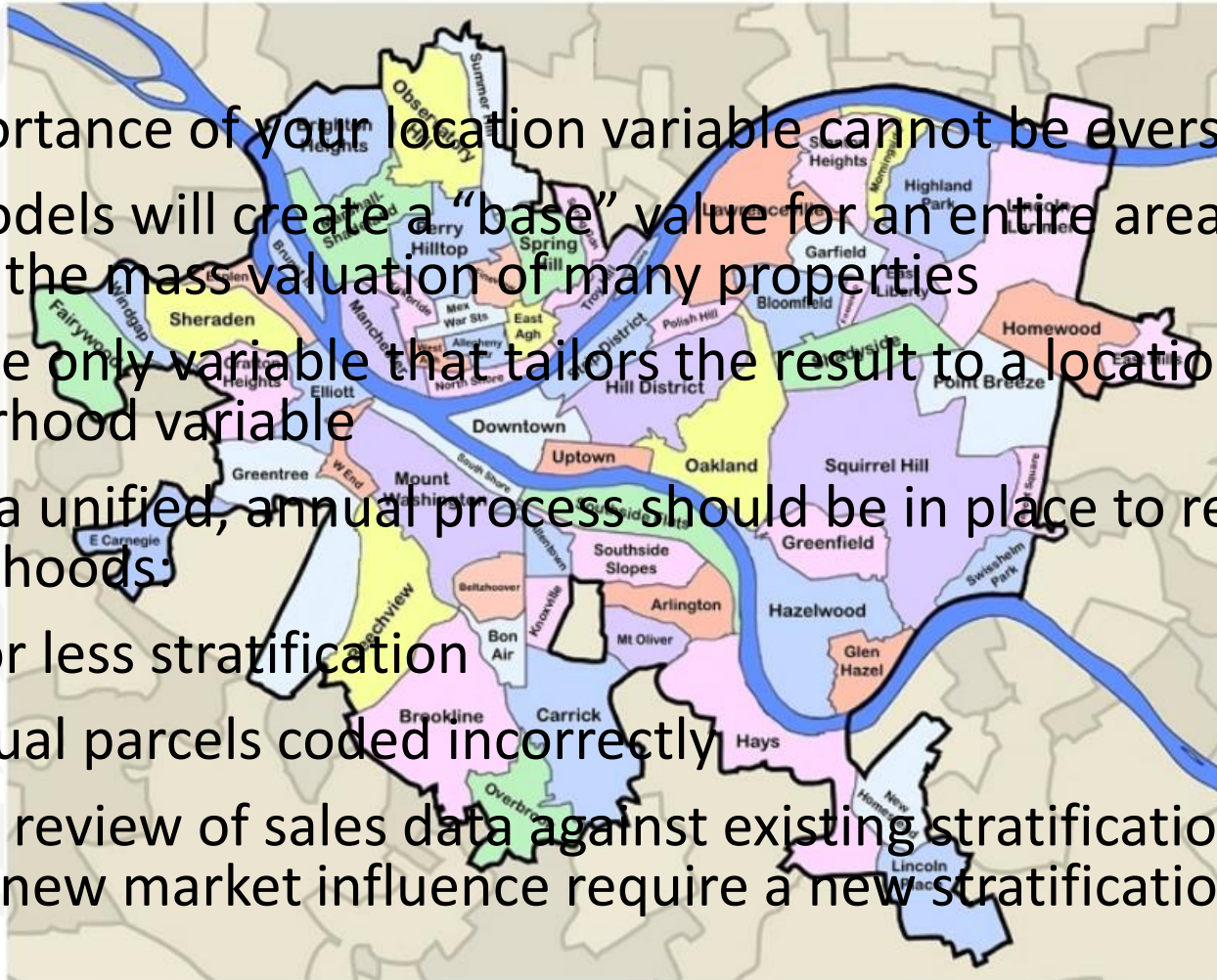
The importance of your location variable cannot be overstated.

Many models will create a “base” value for an entire area, enabling the mass valuation of many properties

Often, the only variable that tailors the result to a location is the Neighborhood variable

As such, a unified, annual process should be in place to review neighborhoods:

- More or less stratification
- Individual parcels coded incorrectly
- Annual review of sales data against existing stratifications – does a new market influence require a new stratification?



Specification of the Model

Once you know the appraisal problem and know the type and scope of data available to you, you can begin to determine the model you would like to use

- Additive Model
- Multiplicative Model
- Hybrid Model
- One of any other types:
 - Adaptive Estimation Procedure
 - Automated Comparable Sales
 - Response Surface Analysis

$$S = b_0 + b_1X_1 + b_2X_2 \dots$$

$$S = b_0 * X_1^{b_1} * X_2^{b_2} \dots$$

$$S = b_0 * (b_1X_1 + X_2^{b_2} \dots)$$

Variables Should only be those that Motivate the market

The big ticket items we recognize as moving the market should be included in your models:

- Locational factors
- Physical Characteristics
- Quality Indicators
- Age

Other characteristics should also be included to the extent they represent supply and demand factors recognized by the market:

- Basements
- Garages
- Patios
- Heating and Cooling
- On and On...

Diminishing Returns on Additional Variables

But be careful! There is a diminishing return on the inclusion of additional variables.

- Fireplaces
- Garage finish
- Various land characteristics

Every additional variable may make the model statistically better but also:

- Make the model more complex and harder to explain
- Introduce the potential for additional error
- Betterment may be insignificant

Beta Coefficients can help you find these insignificant ones

Description of Time Adjustments

Appraisal theory states that adjustments for time are some of the first adjustments you make.

Multiple approaches exist:

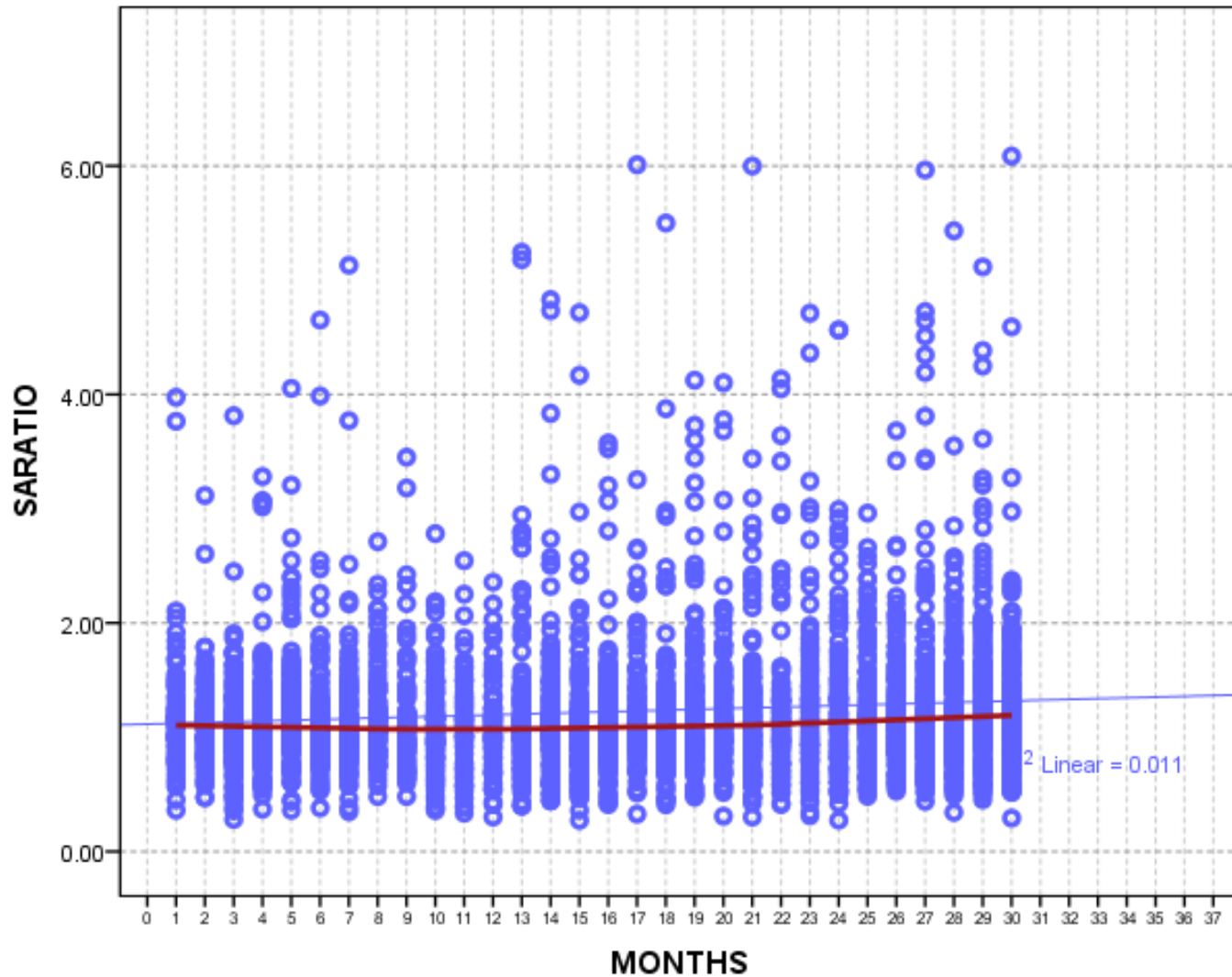
- **Paired Sales** – comparable sales occurring at different periods
- **Resale Analysis** – Sales of the same property
- **Sales Ratio Trend Analysis** – Linear regression of sale price/value over time
- **Multiple Regression Analysis** – time included as one of the variables in a regression analysis

Process of Time Adjustments

Sales Ratio Trend Analysis:

- **Sales are reviewed** so they can be properly coded as valid or invalid
- Time trend **window is selected**, including beginning and end point
- Sales **data is analyzed**, outliers are identified and removed
- Trending **factors are determined** from regression
- Factors are used to create **Time Adjusted Sales Prices** for use in valuation and equalization

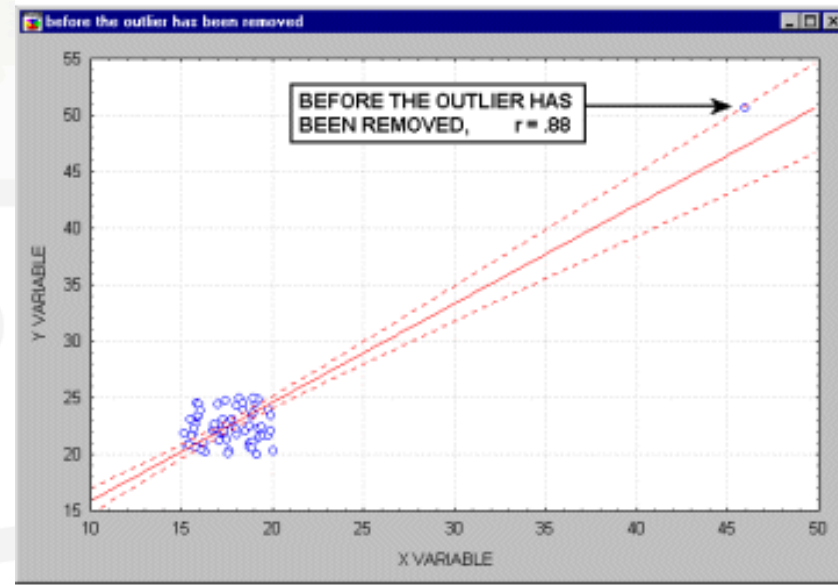
Sample Time Trends



Sales Data Review

Danger! Second occurrence of same cliché: **“GIGO!”**

- Sales affidavits and any other data sources are reviewed prior to acceptance into systems
- Systems are quality controlled to make sure data entry was accurate
- Sales analysis is conducted to make sure sales represent unsold population
- During sales analysis, outliers can be identified and transformed or removed



Sales Must Represent Population

- **Determine clustering** – sales should not be centralized in a single geographic location
- **Sales distribution should match** the distribution of the thing you are valuing (Quality, Age)
- Ideally, you will have an **even span of sales over time**, with no single period over- or under-represented
- To the extent that the sales **do not** represent the population, you may need to adjust what you decide to model

Number of Sales Required for Modeling

- Frequent answers to this question:
- “In general, assessment jurisdictions with a good data base and one hundred or more arms-length residential sales per year can develop effective sales comparison models.” – *Property Appraisal and Assessment Administration*
- Number of sales should be 4 times the number of independent variables – quote from IAAO 311 – Res modeling concepts
- Degrees of Freedom calculations, to get to 95% confidence sample (sales) represents the population

Number of Sales Required for Effective Modeling

- The more the merrier
- Lots
- Mas ventas
- It is a hard question to answer – so much depends on the homogeneity of the sales representing a data element
- 40 sales of widely dispersed sales versus 10 sales of tight, uniform sales?

Calibration of the Model

After specifying the type of model you want to use, you must **determine the values of all the pieces** you want to model – this is **calibration**.

Calibration is achieved by running the data through a model – the ‘**regression**’ in regression modeling

Multiple platforms can be used to run regressions, including Excel, R, SPSS, NCSS, and others

Calibration requires a base understanding of **statistical theory** in order to analyze how well the model is doing

Platform for Regression

The platform you select for modeling will (obviously) affect your ability to use certain processes, conduct certain transformations, use particular analytical tools and ultimately produce the information you need for import into your CAMA system.

You should select a program with:

- Rich analytical tools
- Presentation of results in multiple visualizations
- Robust tools like graphs, statistical summaries

Transformation on Property Characteristics Data

“All data are equal, but some data are more equal than others.”

– *Almost a quote from George Orwell, Animal Farm*

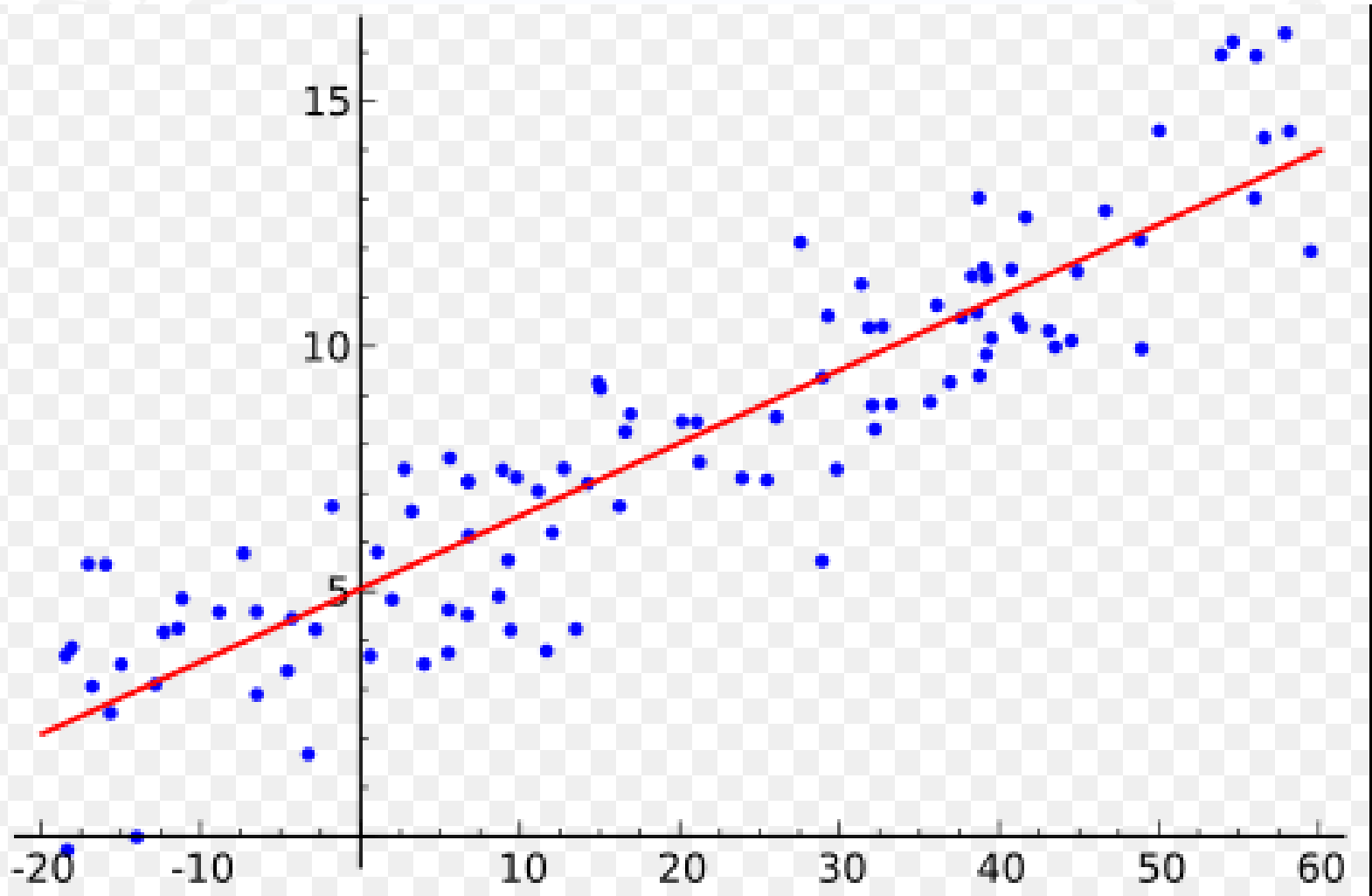
- Some types of data are easily included in models:
 - Scalar information - 1, 2, 3, 4, 5, 3000
 - Binary data - 0,1, 011101000000
- Other types can require transformation:
 - Quality – “Fair”, “Average”, “Good”
 - NBHD Description – “0604 – Lake Mohave Ranchos”

Not exactly a transformation, but new data can also be created from existing data – Golf Course + Ranch = GCRanch

Concepts and Terms used in Regression

- The objective of Regression Analysis is to model the relationship between property characteristics and value.
- The better a model explains the variability in sales prices the better
- Regression uses a “Best Fit” concept – a single mathematical formula can be established that produces the lowest amount of variation between predicted and actual values

Regression model as best fit to data



Concepts and Terms used in Regression

Various statistics help determine how well a model works:

- **Coefficient of Determination (R^2):** the percentage of variance in sales prices explained by the model
- **t-statistic:** a measure of the significance of a variable in explaining differences in the dependent variable
- **Beta Coefficients:** helps compare the relative importance of one variable to another

Examples of Regression Stats

Model Summary

Model	R	R Square ^b	Adjusted R Square	Std. Error of the Estimate
1	.977 ^a	.954	.954	22175.31042
2	.977 ^c	.954		
3	.977 ^d	.954		
4	.977 ^e	.954		
5	.977 ^f	.954		
6	.977 ^g	.954		

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	AGE1SF	-1.125	.064	-.316	-17.445	.000
	AGE2SF	-.181	.032	-.056	-5.735	.000
	AGE3SF	-.042	.132	-.002	-.316	.752
	BASEMENT	-44.083	48.939	-.004	-.901	.368
	RCNADDS	.514	.045	.107	11.457	.000
	SUB01	7616.158	3602.226	.011	2.114	.035
	SUB02	-332.003	1289.670	-.002	-.257	.797
	SUB04	3306.956	1861.718	.015	1.776	.076
	SUB05	2473.867	4765.206	.003	.519	.604
	R1SF	82.275	17.563	.023	4.685	.000
	R2SF	64.759	1.133	.715	57.147	.000
	R3SF	68.662	1.088	.794	63.110	.000
	R4SF	79.790	1.434	.469	55.628	.000
	R5SF	112.500	2.791	.214	40.302	.000

Process of Regression

- Determine your **dependent variable** (Price or Value)
- Determine those **independent variables** important in explaining differences in sales prices (SF, Age, etc)
- Construct the regression statement depending on platform used
- Run the regression
- Identify **outliers**
- Run a clean regression
- Analyze results

Constraining Variables

Statistics is not appraisal!

Your model will initially produce results that are the best at explaining the variability in sales prices in your presented data – ***but often defy appraisal logic.***

- **Negative** values on basements?
- Spas valued at **25K?**
- R2's valued **higher** than R3's?

In cases like this, the modeler must constrain how these variables are represented in the final model – or risk the wrath and ridicule of the constituents

Constraining Variables

However, often times when you constrain a variable you are making the model worse.... *It is no longer the best fit.*

- R-squared stats **decrease**
- Coefficients of Dispersion **increase**

But:

- The model becomes **more explainable**
- The model makes **more appraisal sense**
- The model produces values that are **more defensible**



Avoiding Collinearity

Some variables end up being constrained because they are **collinear** with other variables

- If two variables capture similar market influences, *they may both be modeling the same thing.*
- The model does not know exactly how to distribute the value between the collinear variables, as they may both produce similar explainability in the model.
- Often presents itself in the model as one variable having waaaay too much value, and its counterpart having waaaay too little value
- Example: The model values a spa at 25k, but the PPSF of the R4 home is lower than R3 homes
- Statistical tests exist for determining how collinear variables are, but they are often obvious, and constraining the variables to a reasonable value often fixes the issue.

Model Stability from year to year

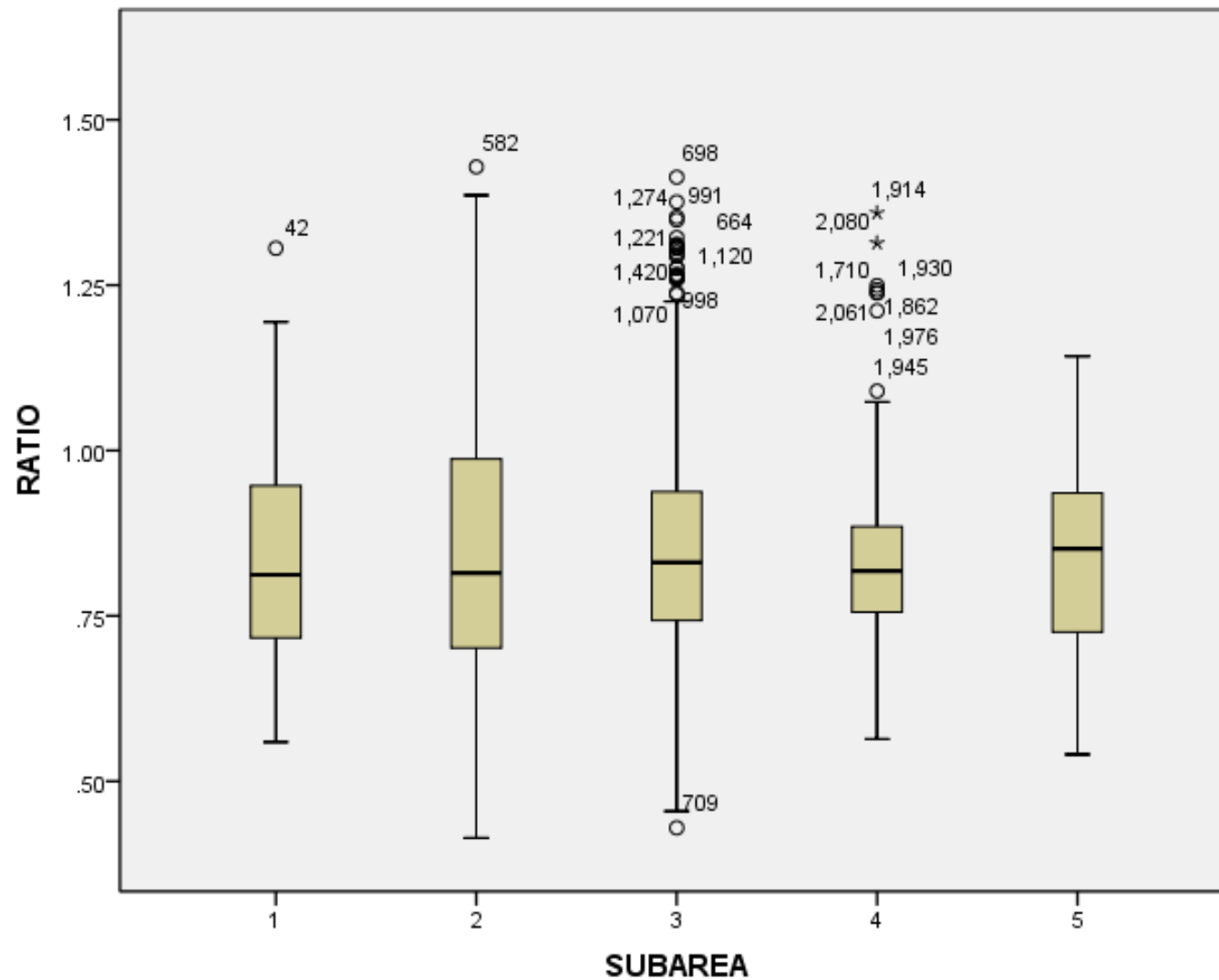
- Regression models are highly agile which also means they can be highly volatile.
- Models should remain stable from one year to the next so that less homogeneous property values do not fluctuate wildly.
- The conscientious modeler will always bounce the new model against the prior year in order to look for large differences in coefficients
- Need to also ensure “expected” variables always have a representation in the annual model.



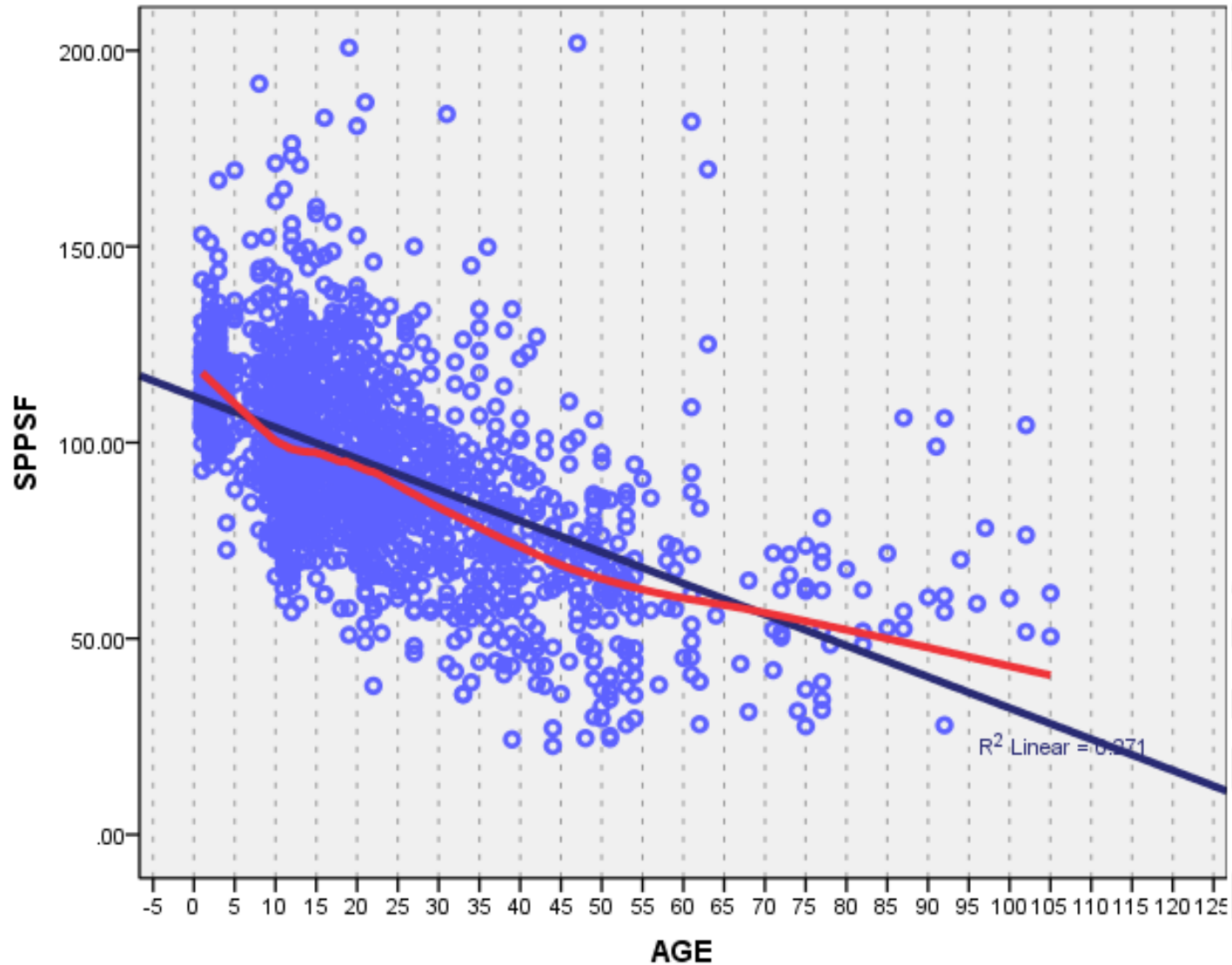
Model Testing and Quality Control

- MSR, COD, PRD are great indicators
 - General Stats against the entire dataset
 - Specific Stats against various stratifications of data
 - By Age
 - By NBHD
 - By Quality
- Boxplots
- Scatter charts
- Outlier reviews
- Visualization of tabular data

Boxplots



Scatter Charts



Compliance versus Appraisal

Appraisal is not equalization and equalization is not appraisal, although plenty of touchpoints exist.

- The sales database used for modeling does not have to be exactly the same as the sales database the DOR will use for equalization
- **BUT IT HELPS WHEN YOU NEED TO ARGUE WITH THE DOR!**
- The modeler needs to walk a fine line between using all the sales that produce the best valuation, and yet also being able to accurately predict equalization results

Compliance versus Appraisal

Communication with the oversight agency is imperative.

- If you run your own time adjustments, consultation with the DOR often results in better results, and both the jurisdiction and the DOR use the same adjustments both for valuation and equalization.
- Insufficient sales may mean an expanded sales range – *but model results should always be reviewed in the context of equalization*
- Reach compliance, **then** go enjoy your **Christmas Vacation.**

Model Refinement

All of the foregoing processes are used to determine what extent the model can be refined in order to produce the lowest CODs and best PRDs while maintaining equilibrium with appraisal concepts

- Statistical review of model fit
- Review of Visualizations
- Review of level and dispersion measures
- Review of outliers
- Review of appraisal feng shui included in final egghead model.

At any point in the process, the model can be reviewed, tweaked and rerun for overall benefit.

- Additional transformations?
- Variable removal or addition?
- Complete Respecification?

Import into CAMA System

Finally, the values must be brought back into the CAMA System.

Heavily dependent on the system

- Input actual mathematical model?
- Input coefficients into pre-structured model in system?
- Bring in model results as whole values (overrides)

Imported values must be tested to make sure the CAMA system has what you intended!

Final thoughts and Questions

“He who fights with monsters should look to it that he himself does not become a monster. When you gaze long into the abyss, the abyss will gaze back into you.”

– Friedrich Nietzsche

Never lose sight of the fact that this is primarily about appraisal, not statistics.

Questions?

Live Demo of Concepts Presented

